Research Paper | Engineering | E-ISSN No : 2454-9916 | Volume : 3 | Issue : 6 | June 2017

RELATIONAL DATABASES WATERMARKING TECHNIQUE BASED ON EMBEDDED PROPORTION

Anuj Kumar Dwivedi [1] | Dr. B. K. Sharma [2] | Dr. A. K. Vyas [3]

[1] Research Scholar, Jodhpur National University, Jodhpur, India.
[2] Professor, Ajay Kumar Garg Engineering College, Ghaziabad, India.
[3] Faculty of E & T, Jodhpur National University, Jodhpur, India.

ABSTRACT

With the rapid growth of web base environment and extensive requirement of databases, the databases owners are need to be care full about to maintain their ownership of the databases. Digital watermarking is the effective solution to protecting the copyright of databases from illegal copying by using the inherent properties of relational databases. In this paper We propose an idea to design an algorithms of watermark embedding and detection by using a predefined parameter, embedded proportion for the purpose of random selection of tuples. We incorporate the use of hash function along with owner id for security reason. In detection process a cutoff criterion is designed to verify watermarked database. It is assumed that the database can tolerate some modifications within a certain limit. Some experiments are conducted over watermark embedding and detection algorithm and the results are analyzed to show the robustness of the proposed scheme by conducting the modifying tuples or attributes attack.

KEYWORDS: ownership protection; cutoff criteria matched value; robustness.

1. Introduction

Copyright protection of owners is becoming more and more necessary due to the rapid growth of the Internet, the wide development of digital multimedia contents, data become easier and easier for distribution. Digital watermarking provides an effective method of protecting digital data from illegal copying, and tempering by embedding a secret code directly into the data. Digital watermarking allows the user to add a layer of protection to the digital media content by identifying copyright ownership and empowered with a tracking capability [1]. Accordingly, it monitors and reports where the user's digital media contents are being used. Now the research of digital watermark technique focuses on the research on relational databases watermark technique. increasing use of databases in applications is creating a need for protecting data copyright in databases [2, 3, 4], and the owners of relational databases worry about their data being pirated and about their ownership protection.

1.1 Technical challenges of database watermarking [6]

There are many differences between the structures of multimedia data and relational databases. Therefore, the watermarking process on relational database is challenged by the following factors:

i. **Few redundant data:** A relational database is made up of tuples, each indicating an independent object. Therefore, watermarks basically have no places to hide or embed whereas multimedia object consists of a large number of bits with considerable redundancy. Thus, the watermark has a large cover in which to hide.

ii. **Out-of-order relational data:** Tuples of a relational database have no fixed location. This makes building a corresponding relative is very difficult in relational databases. However relative spatial/temporal positioning of various pieces of a multimedia object typically does not change.

iii. **Frequent updating:** Insertion, dropping, updating of operation of relational database is very frequent. Without malicious intention, users often casually drop some tuples or attributes. On the other hand, the pirate can add or substitute the tuples and attributes whereas, multimedia objects typically remain intact; portions of an object cannot be dropped or replaced arbitrarily without causing perceptual changes in the object. Because of these differences, techniques developed for multimedia data cannot be directly used for watermarking relations.

1.2 Requirements of database water marking

Water marking database has unique requirements that differ from those required for watermarking digital image and audio systems. The watermarked database must maintain the following properties [6]:

i. **Usability:** The amount of change in the database caused by the water marking process should not result in degrading the database and making it useless. The amount of allowable change differs from one database to another, depending on the nature of stored records.

ii. **Robustness:** Watermarks embedded in the database should be robust against attacks to erase them. That is, the database watermarking algorithm must be developed in such a way to make it difficult for an adversary to remove or alter the watermark beyond detection without destroying usability of the database.

In this paper, we proposed an idea of protecting piracy of relational databases through embedding watermark based on the embedded proportion and importance of attribute content.

The rest of the paper is organized as follows.

Section 2 simply describes the previous research schemes of this field. Section 3 specifies the watermarking algorithms based on embedded proportion and content importance. Section 4 gives a formal interpretation of the algorithms through the optimistic probability. Section 5 conducted some experiments and analyzed its robustness. And section 6 draw some conclusions.

2. Previous research schemes of Database Watermarking

In 2002, Agrawal and Kiernan first proposed a robust watermark for databases in. This method marks only numeric attributes and proposed the idea of watermarking using least significant bits (LSB). They do not account for multibit watermarks which make their technique vulnerable against simple attacks, for example, shifting of only one least significant bit results in loss of watermark [2]. It has some flaws that the embedded marks should be closely related to the primary key attribute of relational databases, the primary key attribute value could not be modified or replaced, or else, the scheme would have no meaning.

In 2008, Sun et al. introduced another technique for inserting an image into the database as watermark information. In this method, they converted one or more images into flow of bits. They used hash value of database tuple to find the location of each pixel and marked bit. They considered mod of hash value and watermark's length. If someone takes large image as watermark information, then length of watermark increases. And this method cannot insert all the pixels into the database. Therefore, this method is not efficient for small databases [5].

In 2011, Min, Li and Wenyue, Zhao proposed an asymmetric watermarking scheme that employed the digital signature technology. A message can be signed by the owner with a private key. Anyone can verify this signature using the corresponding public key. Signature cannot be forged, that is signed message is indeed from the private key holder. [6]

In 2012, U. Pratap et al. propose, a new technique of database which based on inserting the bits of a binary image in relational database. The proposed technique also minimizes the variation in watermarked database. Experimental results justify the feasibility of the proposed technique and its robustness against common database attacks [6].

We have already published a survey paper by summarizing and analyzing of above mentioned previous research work and now Combined with the merits of above mentioned related research work, we proposed an effective watermarking relational databases technique scheme based on embedded proportion and opti-

mistic probability without using any specific database key.

3. Proposed watermarking algorithm

In our scheme we select attributes one by one and watermarking of it decided on the basis of random proportion it needs a cutoff value which indicate that the certain errors are tolerated for some numerical attributes of relational databases. we can extract the bits of an attribute within a tuple as a mark, and then embed the mark into another weaker attribute within the tuple, then the tuple is called the matched tuple. The ratio of the matched tuples in relational databases and the detection threshold value (mv) used to validate the existence of digital watermark (that is the copyright of relational databases). According to the idea, the watermark embedding and detection algorithms are proposed in section 3.1 and section 3.2 respectively. And some notations which known only by the owners used in the algorithms are described in table 1.

Table 1 Abbreviations

Ai:	the attribute from which embedded portion of watermark is extracted.
A:	the weaker attribute of the database where the mark is to be embedded.
L:	used to select the number of bits of A1. The length of the extracted characteristic bits
S:	used to select the number of L bits of Ai. the L bits are chosen from the Sth position of the starting of the binary value of Ai
d:	hash value with owner id for security purpose.
e:	the embedded proportion of relational databases (0<e<=1)
mv:	match value used to detect watermark
tw:	total watermarked tuple
op:	optimistic probability selected on the basis of some facts.

3.1 Watermark Embedding Algorithm

The watermark embedding algorithm in details is described in Algorithm 1. Here RDB denotes original relational database. RDBW denotes watermarked relational databases. The parameters L, S and e, op and OID are known only to the owner of the databases. In algorithm 1, The function random (0..1) is used to generate a sequence of equally distributed random number In the range of 0 and 1. If the generated random number is. smaller than the embedded proportion e, and the value of numeric attribute Ai is not null, we can watermark the tuple. As step 8,9 and 10 if random proportion criteria meet then tuple will be marked otherwise not, thus the method of choosing tuples for marking is random. Now to improve security attribute value and owner id hashed together and we convert the result into binary string and extract a predefined length of string for embedding in a relatively weaker attribute to keep database usability intact.

Algorithm1:
1) Input: (RDB);
2) tw =0;
3) While (RDB ≠EOF) do
4) for each tuple ∈ RDB
5) If random(0..1) is less than e and Ai.value within the current tuple is not null then
6) d = H (OID, Ai.value)
7) convert the integer value d to a binary sequence and note the value of sequence in Abi, i is an integer;
8) extract the L bits from Abi, Wi = BString(Abi, S, L);
9) convert Wi to IWi (integer value)
10) embedded the IWi into the last of A.value;
11) tw++
12) end;
13) RDB.next;
14) End;

Figure 1: embedding algorithm

.

3.2 Watermark Detection Algorithm

The watermark detection algorithm is described in algorithm 2. Here, we verify the suspicious RDB. The parameters L, S, e and OID have the same value used for the watermark embedding. In algorithm 2, step 1 denotes that the variable totalSRDB equals the number of tuples in suspicious RDB. And the problem of watermark detection is judged by the parameter mv denotes the match value calculated during detection process.

Algorithm2:
1) Input: (suspicious RDB);
2) totalSRDB = total tuples (suspicious RDB);
3) mv = 0;
4) While (RDB ≠EOF) do
5) for each tuple ∈ SRDB
6) if Ai.value within the current tuple is not null then
7) d = H(OID, Ai.value)
8) convert the integer value d to a binary sequence and note the value of sequence in Abi, i is an integer;
9) extract the L bits from Abi, Wi = BString(Abi, S, L);
10) convert string L to Iwi
11) if IWi matched with the last of A.value then
12) mv ++;
13) end;
14) SRDB.next;
15) End;
16) if tw/mv > (e + op) / 2 then
return true; // Owner retrieves his watermark from the suspected relation and R is recovered successfully
17) has watermark
18) else
19) return false; // suspected relation RDB cannot be recovered
20) no watermark

Figure2: detection algorithm

embedded proportion e chosen arbitrarily. The function BString(Abi, S, L) is used to extract a string as the watermark bits from Abi, the length of the section string is L and it is extracted from the Sth position of the string Abi., and noted as Wi. If the length of the string Abi is less than L, the empty position of Abi will be filled with '0'. For example, BString('11001010',5, 3), the extract function returns Wi equals '101'. And after that convert extracted bits to integer and embed with the identified weaker attribute.

4. Algorithm Analysis

When the detected ratio of the matched tuples and matched value (mv) is greater than the cutoff value then we can predict that the watermark exists. If the embedded proportion produced by random function is greater than e we do not think that the watermarking tuples exist in relation. According to the above watermarking algorithms, the marked tuples are randomly selected which is related to the embedded proportion. When the embedded proportion e will always greater than the optimistic probability op (chosen after observing some experimental results), the number of watermarking tuples should lies in the range of 0 and e × N , and the number of the matched tuples should lies in the range of op × N and (e + op)× N . So the detection probability is higher when ratio of tw/mv is greater than (e + op) / 2. When the detection ratio of the matched tuples is greater than the selected detection cutoff value, we can then confirm that the watermark exists in relational databases.

4.1 Robustness analysis

Robustness is the basic requirement of watermark technique. In this section we show that our watermarking scheme persist against moderation up to a certain limit There are four typical attacked modes include:

i. **Subset deletion attack:** In this type of attack, the attacker may take a subset of the tuples of the watermarked database and hope that the watermark will be removed.

ii. **Subset addition attack:** In this type of attack, the attacker adds a set of tuples to the original database. This is one of the most difficult attacks to defeat. The attacker may add some tuples to the watermarked table.

iii. **Subset alteration attack:** In this type of attack, the attacker alters the tuples of the database through operations such as linear transformation. The attacker hopes by doing so to erase the watermark from the database.

iv. **Subset selection attack:** In this type of attack, the attacker randomly selects and uses a subset of the original database that might still provide value for its intended purpose. The attacker hopes by doing so that the selected subset will not contain the watermark [4].

For our watermarking scheme, we considering subset alteration attack assuming that optimistic probability is 0.1. In this experiments, we take size of relational databases be 10,000, for the different embedded proportion, from table 2, we can see that the scheme can robust against subset modified attack within the certain limit.

Table 2 Detection of watermark under the subset modification attack

Embedded Proportion	Modified Proportion	Ratio of matched tuples	Existence of watermark
.15	0	0.1532	yes
	.05	0.1510	yes
	.25	0.1475	yes
	.5	0.1295	yes
	.75	0.1025	no
.25	0	0.2515	yes
	.05	0.2422	yes
	.25	0.2041	yes
	.5	0.1831	yes
	.75	0.1581	no

.50	0	0.4115	yes
	.05	0.3987	yes
	.25	0.3512	yes
	.5	0.2716	no
	.75	0.1881	no
.75	0	0.8514	yes
	.05	0.5215	yes
	.25	0.3546	no
	.5	0.2154	no
	.75	0.1181	no

When the embedded proportion equals 0.50 and the modified proportion is 0.50, the watermark cannot be detected from relational databases. However, when the embedded proportion is closer to 1, the watermark can be detected from relational databases with higher embedded proportion only when the modified proportion is very less (0.05). Thus table 2 shows that capability of watermarking scheme against the subset modification attack is depend upon the embedded proportion. We should select the appropriate embedded proportion to stand against the subset modified attack.

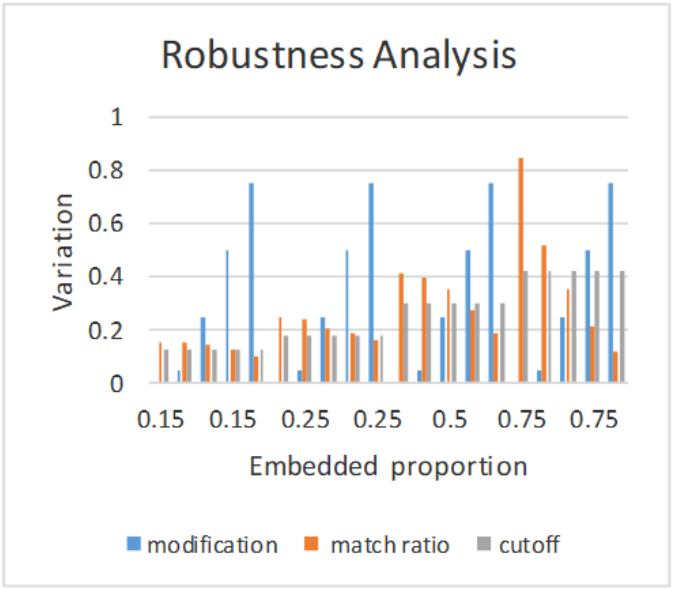

Figure 3 detection of watermark after subset modification

From the figure 3 it can be easy to see that as the embedded proportion increases the tolerance limit of modification is decreases. When the modification value increases then match ratio decreases and after a certain limit match ratio remain below from the cutoff value. thus for the high modification rate with big embedded proportion usability of database will not sustain but for smaller embedded proportion(e) our scheme is robust.

5. Conclusions and future scope:

In this paper, we have proposed a water mark technique that is based on a random embedded proportion value and optimistic probability. We involve owner id for the purpose of improving security. In this scheme we have not use any specific key value from the database so it will difficult to correlate marked tuples and attribute. Attacker cannot guess any order in the marked tuples due to randomness. Based on the experimental robustness analysis against malicious attacks, the results show that the proposed scheme of watermarking relational databases is robust. The experiments show that the amount of errors introduced to the relation can be tolerated or neglected up to a certain limit. Our future research will be focus towards embedding of watermark without compromising the usability of original database without using probability because most of the detection algorithms are probability based. And also increase the level of security attacks and try to prove that our scheme resilience against several different type of attacks

REFERENCES

1. Saraju P. Mohanty , "Digital Watermarking : A Tutorial Review" Indian Institute of Science, Bangalore, 1999
2. R. Agrawal, J. Kiernan. "Watermarking Relational Databases", In: Proceeding of the 28th VLDB Conference. Hong Kong, 2002: 155-166.
3. G.H. Gamal, M.Z. Rashad and M.A. Mohamed "A Simple Watermark Technique for Relational Databa" Mansoura Journal for Computer Science and Information Systems Vol. 4, No.4, Jan2008.
4. Raju Halder, Shantanu Pal, Agostino Cortesi "Watermarking Techniques for Relational Databases: Survey, Classification and Comparison" Journal of Universal Computer Science, Vol. 16, no.21 2010, pp.3165-3190
5. Min, Li, Wenyue, Zhao, "An Asymmetric Watermarking Scheme for Relational Database", Communication Software and Networks (ICCSN), IEEE 3rd International Conference. 2011,pp.180-184
6. Udai Pratap Rao a, Dhiren R. Patel a, Punitkumar M. Vikani, "Relational Database Watermarking for Ownership Protection 2nd International Conference on Communication, Computing & Security [ICCCS-2012] Science Direct pp.988-995.